

特 許 協 力 条 約

PCT

国際予備審査報告

(法第12条、法施行規則第56条)
〔PCT36条及びPCT規則70〕

REC'D 19 SEP 2003

WIPO

PCT

出願人又は代理人 の書類記号 663491	今後の手続きについては、国際予備審査報告の送付通知(様式PCT/ IPEA/416)を参照すること。	
国際出願番号 PCT/JPO2/13053	国際出願日 (日.月.年) 13.12.02	優先日 (日.月.年) 16.01.02
国際特許分類 (IPC) Int. Cl' G10L15/06, G10L15/18		
出願人 (氏名又は名称) シャープ株式会社		

1. 国際予備審査機関が作成したこの国際予備審査報告を法施行規則第57条 (PCT36条) の規定に従い送付する。

2. この国際予備審査報告は、この表紙を含めて全部で 3 ページからなる。

☒ この国際予備審査報告には、附属書類、つまり補正されて、この報告の基礎とされた及び/又はこの国際予備審査機関に対してした訂正を含む明細書、請求の範囲及び/又は図面も添付されている。
(PCT規則70.16及びPCT実施細則第607号参照)
この附属書類は、全部で 6 ページである。

3. この国際予備審査報告は、次の内容を含む。

I ☒ 国際予備審査報告の基礎

II ☐ 優先権

III ☐ 新規性、進歩性又は産業上の利用可能性についての国際予備審査報告の不作成

IV ☐ 発明の単一性の欠如

V ☒ PCT35条(2)に規定する新規性、進歩性又は産業上の利用可能性についての見解、それを裏付けるための文献及び説明

VI ☐ ある種の引用文献

VII ☐ 国際出願の不備

VIII ☐ 国際出願に対する意見

国際予備審査の請求書を受理した日
07.08.03

国際予備審査報告を作成した日
01.09.03

名称及びあて先
日本国特許庁 (IPEA/JP)
郵便番号100-8915
東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

5C

9379

樫本 剛



電話番号 03-3581-1101 内線 3541

I. 国際予備審査報告の基礎

1. この国際予備審査報告は下記の出願書類に基づいて作成された。(法第6条(PCT14条)の規定に基づく命令に
応答するために提出された差し替え用紙は、この報告書において「出願時」とし、本報告書には添付しない。
PCT規則70.16, 70.17)

☐ 出願時の国際出願書類

☒ 明細書 第 1-2, 7-12 ページ、 出願時に提出されたもの
明細書 第 3-6 ページ、 国際予備審査の請求書と共に提出されたもの
明細書 第 _____ ページ、 _____ 付の書簡と共に提出されたもの

☒ 請求の範囲 第 2-4, 8 項、 出願時に提出されたもの
請求の範囲 第 _____ 項、 PCT19条の規定に基づき補正されたもの
請求の範囲 第 1, 5-7 項、 国際予備審査の請求書と共に提出されたもの
請求の範囲 第 _____ 項、 _____ 付の書簡と共に提出されたもの

☒ 図面 第 1-9 ページ/図、 出願時に提出されたもの
図面 第 _____ ページ/図、 国際予備審査の請求書と共に提出されたもの
図面 第 _____ ページ/図、 _____ 付の書簡と共に提出されたもの

☐ 明細書の配列表の部分 第 _____ ページ、 出願時に提出されたもの
明細書の配列表の部分 第 _____ ページ、 国際予備審査の請求書と共に提出されたもの
明細書の配列表の部分 第 _____ ページ、 _____ 付の書簡と共に提出されたもの

2. 上記の出願書類の言語は、下記に示す場合を除くほか、この国際出願の言語である。

上記の書類は、下記の言語である _____ 語である。

- ☐ 国際調査のために提出されたPCT規則23.1(b)にいう翻訳文の言語
☐ PCT規則48.3(b)にいう国際公開の言語
☐ 国際予備審査のために提出されたPCT規則55.2または55.3にいう翻訳文の言語

3. この国際出願は、ヌクレオチド又はアミノ酸配列を含んでおり、次の配列表に基づき国際予備審査報告を行った。

- ☐ この国際出願に含まれる書面による配列表
☐ この国際出願と共に提出された磁気ディスクによる配列表
☐ 出願後に、この国際予備審査(または調査)機関に提出された書面による配列表
☐ 出願後に、この国際予備審査(または調査)機関に提出された磁気ディスクによる配列表
☐ 出願後に提出した書面による配列表が出願時における国際出願の開示の範囲を超える事項を含まない旨の陳述書の提出があった
☐ 書面による配列表に記載した配列と磁気ディスクによる配列表に記録した配列が同一である旨の陳述書の提出があった。

4. 補正により、下記の書類が削除された。

☐ 明細書 第 _____ ページ
☐ 請求の範囲 第 _____ 項
☐ 図面 図面の第 _____ ページ/図

5. ☐ この国際予備審査報告は、補充欄に示したように、補正が出願時における開示の範囲を越えてされたものと認められるので、その補正がされなかったものとして作成した。(PCT規則70.2(c) この補正を含む差し替え用紙は上記1.における判断の際に考慮しなければならず、本報告に添付する。)

V. 新規性、進歩性又は産業上の利用可能性についての法第12条（PCT35条(2)）に定める見解、それを裏付ける文献及び説明

1. 見解

新規性 (N)

請求の範囲 1-8 有
請求の範囲 無

進歩性 (IS)

請求の範囲 1-8 有
請求の範囲 無

産業上の利用可能性 (IA)

請求の範囲 1-8 有
請求の範囲 無

2. 文献及び説明 (PCT規則70.7)

請求の範囲 1-8

文献1：李，河原，竹田，鹿野，「Phonetic Tied-Mixture モデルを用いた大語彙連続音声認識」，電子情報通信学会技術研究報告「言語理解とコミュニケーション」，1999.12.20，Vol.99，No.523，NLC99-32，Pages 43-48

には、当該技術分野の一般的技術水準を示す文献として、複数の環境依存音響モデルの状態系列をまとめたものが記載されているが、前記状態系列を木構造化した点に関しては、国際調査報告で列記した文献のいずれにも、記載も示唆もされていない。

これに対して、特開平 1 1 - 4 5 0 9 7 号公報に開示された連続音声認識方式においては、前後の単語に依存せずに決まる音響モデル系列を認識単語として記述した認識単語辞書と、単語境界において前後の単語に依存して記述した単語間単語辞書を用いて照合を行うようにしている。したがって、単語境界にも音素環境依存の音響モデルを用いることによって精度を確保しながら、大語彙の場合でも単語境界での処理量の増大を抑えることができるのである。しかしながら、一般に、単語のスコアや境界はそれ以前の単語の影響を受けるので、複数の認識単語が単語間単語を共有すると、図 9 A に示すように認識単語 “k;o:k” および “s;o:k” と単語間単語 “o” との境界の履歴が考慮されないので、図 9 B に示すように単語の境界履歴を考慮した場合に比して、性能の低下を招く恐れがある。また、例えば助詞の “を (/o/ と発声)” 等のように、認識単語辞書と単語間単語辞書とに分割することができない単語については開示されていない。

発明の開示

そこで、この発明の目的は、単語境界にも音素環境依存音響モデルを用いて精度を確保しつつ、大語彙の連続音声認識時にも単語境界での処理量の増大を抑えることができる連続音声認識装置および連続音声認識方法、連続音声認識プログラム、並びに、連続音声認識プログラムを記録したプログラム記録媒体を提供することにある。

上記目的を達成するため、この発明は、隣接するサブワードに依存して決定されるサブワードを認識単位とすると共に、サブワード環境に依存する環境依存音響モデルを用いて、連続的に発声された入力音声进行を認識する連続音声認識装置であって、語彙中の各単語が、サブワードのネットワークあるいはサブワードの木構造として格納された単語辞書と、単語間の接続情報を表す言語モデルが格納された言語モデル格納部と、上記環境依存音響モデルが、当該環境依存音響モデルの状態系列のうち、複数のサブワードモデルの状態系列をまとめて木構造化して成るサブワード状態木として格納されている環境依存音響モデル格納部と、上記環境依存音響モデルであるサブワード状態木、上記単語辞書および言語モデルを参照して上記サブワードの仮説を展開すると共に、入力された音声の特徴パラメ

一タと上記展開された仮説との照合を行い、単語の終端に該当する仮説に関する単語、累積スコア及び始端開始フレームを含む単語情報を出力する照合部と、上記単語情報に対する探索を行って認識結果を生成する探索部を備えたことを特徴としている。

- 5 上記構成によれば、サブワード環境に依存する環境依存音響モデルを木構造化したサブワード状態木、単語辞書および言語モデルを参照して、サブワードの仮説を展開するようにしている。したがって、次に続く単語の先頭サブワードに関係無く1つの仮説を展開すればよく、全仮説における状態の総数を削減することができる。すなわち、仮説の展開処理量を大幅に削減でき、単語内および単語境界に関係なく、仮説の展開が容易になるのである。さらに、照合部によって、上記音響分析部からの特徴パラメータ系列と上記展開された仮説との照合を行う際における照合処理量が大幅に削減される。
- 10

- また、1実施例では、上記発明の連続音声認識装置において、上記環境依存音響モデル格納部に格納されている環境依存音響モデルは、中心サブワードが前後のサブワードに依存する環境依存音響モデルのうち、先行サブワードおよび中心サブワードが同じサブワードモデルの状態系列を木構造化したサブワード状態木である。
- 15

- この実施例によれば、先行サブワードおよび中心サブワードが同じサブワードモデルの状態系列を木構造化したサブワード状態木を用いて、上記仮説を展開している。したがって、次の仮説を展開する場合には、終端仮説における中心サブワードのみに注目して対応する先行サブワードを有するサブワード状態木を展開すればよい。つまり、後続サブワードが複数あってもより少ない仮説を展開すればよく、仮説の展開が容易である。
- 20

- また、1実施例では、上記発明の連続音声認識装置において、上記環境依存音響モデルは、複数のサブワードモデルで状態を共有している状態共有モデルである。
- 25

 この実施例によれば、複数のサブワードモデルによって状態を共有することによって、木構造化した際に共有している状態を一つにまとめることができ、ノード数を削減することができる。したがって、上記照合部による照合時における処

理量が大幅に削減される。

また、1実施例では、上記発明の連続音声認識装置において、上記照合部は、上記サブワード状態木を参照して仮説を展開する際に、上記単語辞書および言語モデルから得られる接続可能なサブワード情報を用いて、上記仮説であるサブワード状態木を構成する状態のうち、互いに接続可能な状態にフラグを付すようになっている。

この実施例によれば、上記展開された仮説を構成するサブワード状態木の状態のうち、互いに接続可能な状態のみにフラグを付けるようにしたので、上記照合の際にビタビ計算を行う必要がある状態が限定されて、照合処理量が更に削減される。

また、1実施例では、上記発明の連続音声認識装置において、上記照合部は、上記照合を行う際に、上記特徴パラメータに基づいて上記展開された仮説のスコアを算出すると共に、このスコアの閾値あるいは仮説数を含む基準に従って上記仮説の枝刈りを行うようになっている。

この実施例によれば、上記照合時に仮説の枝刈りを行うので、単語となる可能性が低い仮説が削除されて、以後の照合処理量が大幅に削減される。

また、この発明は、隣接するサブワードに依存して決定されるサブワードを認識単位とすると共に、サブワード環境に依存する環境依存音響モデルを用いて、連続的に発声された入力音声进行を認識する連続音声認識方法であって、照合部によって、上記環境依存音響モデルの状態系列を木構造化して成るサブワード状態木、語彙中の各単語がサブワードのネットワークあるいはサブワードの木構造として記述された上記単語辞書、および、単語間の接続情報を表す言語モデルを参照して、上記サブワードの仮説を展開すると共に、入力された音声の特徴パラメータと上記展開された仮説との照合を行って、単語の終端に該当する仮説に関する単語、累積スコアおよび始端開始フレームを含む単語情報を生成し、探索部によって、上記単語情報に対する探索を行って認識結果を生成することを特徴としている。

上記構成によれば、上記発明の連続音声認識装置の場合と同様に、環境依存音響モデルを木構造化したサブワード状態木を参照して仮説を展開するので、次に続く単語の先頭サブワードに関係無く1つの仮説を展開すればよく、単語内およ

び単語境界に関係なく仮説の展開が容易になるのである。さらに、特徴パラメータ系列と上記展開された仮説との照合を行う際における照合処理量が大幅に削減される。

また、この発明の連続音声認識プログラムは、コンピュータを、上記発明の連続音声認識装置における単語辞書、言語モデル格納部、環境依存音響モデル格納部、照合部および探索部として機能させることを特徴としている。

上記構成によれば、上記発明の連続音声認識装置の場合と同様に、次に続く単語の先頭サブワードに関係無く1つの仮説を展開すればよく、単語内および単語境界に関係なく仮説の展開が容易になる。さらに、特徴パラメータ系列と上記展開された仮説との照合を行う際における照合処理量が大幅に削減される。

また、この発明のプログラム記録媒体は、上記発明の連続音声認識プログラムが記録されたことを特徴としている。

上記構成によれば、上記発明の連続音声認識装置の場合と同様に、次に続く単語の先頭サブワードに関係無く1つの仮説を展開すればよく、単語内および単語境界に関係なく仮説の展開が容易になる。さらに、特徴パラメータ系列と上記展開された仮説との照合を行う際における照合処理量が大幅に削減される。

図面の簡単な説明

図1は、この発明の連続音声認識装置におけるブロック図である。

図2A、図2Bは、音素環境依存音響モデルの説明図である。

図3は、図1における単語辞書の説明図である。

図4は、言語モデルの説明図である。

図5A、図5Bは、図1における前向き照合部による仮説の展開の説明図である。

図6は、上記前向き照合部によって実行される前向き照合処理動作のフローチャートである。

請 求 の 範 囲

1. (補正後) 隣接するサブワードに依存して決定されるサブワードを認識単位とすると共に、サブワード環境に依存する環境依存音響モデルを用いて、連続的に発声された入力音声認識する連続音声認識装置であって、

語彙中の各単語が、サブワードのネットワークあるいはサブワードの木構造として格納された単語辞書(4)と、

単語間の接続情報を表す言語モデルが格納された言語モデル格納部(5)と、

上記環境依存音響モデルが、当該環境依存音響モデルの状態系列のうち、複数のサブワードモデルの状態系列をまとめて木構造化して成るサブワード状態木として格納されている環境依存音響モデル格納部(3)と、

上記環境依存音響モデルであるサブワード状態木、単語辞書(4)および言語モデルを参照して上記サブワードの仮説を展開すると共に、入力された音声の特徴パラメータと上記展開された仮説との照合を行い、単語の終端に該当する仮説に関する単語、累積スコアおよび始端開始フレームを含む単語情報を出力する照合部(2)と、

上記単語情報に対する探索を行って認識結果を生成する探索部(8)を備えたことを特徴とする連続音声認識装置。

2. 請求項1に記載の連続音声認識装置において、

上記環境依存音響モデル格納部(3)に格納されている環境依存音響モデルは、中心サブワードが前後のサブワードに依存する環境依存音響モデルのうち、先行サブワードおよび中心サブワードが同じサブワードモデルの状態系列を木構造化したサブワード状態木であることを特徴とする連続音声認識装置。

3. 請求項2に記載の連続音声認識装置において、

上記環境依存音響モデルは、複数のサブワードモデルで状態を共有している状態共有モデルであることを特徴とする連続音声認識装置。

4. 請求項 1 に記載の連続音声認識装置において、

上記照合部(2)は、上記サブワード状態木を参照して仮説を展開する際に、上記単語辞書(4)および言語モデルから得られる接続可能なサブワード情報を用いて、上記仮説であるサブワード状態木を構成する状態のうち、互いに接続可能な状態にフラグを付すようになっていることを特徴とする連続音声認識装置。

5. (補正後) 請求項 1 に記載の連続音声認識装置において、

上記照合部(2)は、上記照合を行う際に、上記特徴パラメータに基づいて上記展開された仮説のスコアを算出すると共に、このスコアの閾値あるいは仮説数を含む基準に従って上記仮説の枝刈りを行うようになっていることを特徴とする連続音声認識装置。

6. (補正後) 隣接するサブワードに依存して決定されるサブワードを認識単位とすると共に、サブワード環境に依存する環境依存音響モデルを用いて、連続的に発声された入力音声进行を認識する連続音声認識方法であって、

照合部によって、上記環境依存音響モデルの状態系列を木構造化して成るサブワード状態木、語彙中の各単語がサブワードのネットワークあるいはサブワードの木構造として記述された上記単語辞書、および、単語間の接続情報を表す言語モデルを参照して、上記サブワードの仮説を展開すると共に、入力された音声の特徴パラメータと上記展開された仮説との照合を行って、単語の終端に該当する仮説に関する単語、累積スコアおよび始端開始フレームを含む単語情報を生成し、探索部によって、上記単語情報に対する探索を行って認識結果を生成することを特徴とする連続音声認識方法。

7. (補正後) コンピュータを、請求項 1 に記載の単語辞書(4)、言語モデル格納部(5)、環境依存音響モデル格納部(3)、照合部(2)および探索部(8)と